MULTIPLE BIRTH DISCRIMINATION

by

Seymour Geisser*
University of Minnesota
Technical Report No. 168

January 1972

*This work was supported in part by an NIH Grant.

0. Introduction.

A statistical model for multiple birth discrimination is presented. Part of the model was first presented for like sexed twins based on univariate normal assumptions by Richter and Geisser (1960). They also obtained methods for like sexed triplets and quadruplets. The model is now extended to multivariate normal assumptions for like sexed t-tuplets where t is arbitrary. In addition the procedure is improved by obtaining relative weights for the various cases by devising a simple but plausible model from which the weights are derived. A finer discrimination is also obtained in that the new procedure identifies particular individuals with particular eggs. A further significant feature of the model is that the whole discriminatory procedure for the t-tuplet case depends only on the parameters involved in the twin situation. In what follows we shall first deal with the twin case and then the arbitrary t-tuplet case.

1. The Twin Case.

Analagously as in Richter and Geisser (1961) we let $R_1, R_2, \ldots$ now be a sequence of p-dimensional independent random variables such that R_i is $N(\mu_i, \Sigma_W)$. Further let μ_i be an observation on a random variable M which is $N(\mu, \Sigma_B)$. Let x_1, x_2 be p-dimensional observations on a pair of like sexed twins from the same mother. Then x_1, x_2 are interpreted as observations on R_i, R_j respectively. If i=j, the pair are monozygotic (one egg) twins; if $i \neq j$ the pair are dizygotic (two egg) twins. Assume that Σ_W , the within-egg covariance matrix is constant for all mothers. Hence $x_2 - x_1$ is an observation from a $N(0, 2\Sigma_W)$ population if x_1 and x_2 are each independent observations on R_i (the same egg). Now suppose x_1, x_2 are observations on R_i and R_j (different eggs) respectively. Then

x_2-x_1 is $N(\mu_i-\mu_j, 2\Sigma_W)$ given $\mu_i-\mu_j$. Assuming μ_i and μ_j are independent, then $\mu_i-\mu_j$ is distributed as $N(0,2\Sigma_B)$ and further x_2-x_1 is unconditionally distributed as $N(0,2\Sigma_W+2\Sigma_B)$.

Hence the posterior probability that a future twin pair $z=x_2-x_1$, based on the p characteristics is dizygotic is $\varphi_d(z)/(\varphi_d(z)+\gamma\varphi_m(z))$ where φ_d and φ_m represent the density of a $N(0,2\Sigma_W+2\Sigma_B)$ and a $N(0,2\Sigma_W)$ variable respectively while γ is the relative frequency of monozygotic twins to dizygotic like sexed twins in the population from which the new twin pair has been drawn.

2. Arbitrary Number of Offspring.

Now suppose that a birth gives rise to t offspring, $x_1, \ldots, x_t$. Further let $z_i=x_t-x_i$, i=1, ..., t-1 . Since $x_1, \ldots x_t$ are observations assumed normal and independent, the joint set $z_1, \ldots z_{t-1}$ is multivariate normal p(t-1) dimensional, conditional on $\Delta'=(\Delta_1', \ldots, \Delta_{t-1}')$ where $\Delta_i=\mu_t-\mu_i$, i=1 , ..., t-1 . Clearly

$$(2.1)\qquad \begin{cases} E(z_i \mid \Delta)=\Delta_i \\ Cov(z_i \mid \Delta)=2\Sigma_W \\ Cov(z_j, z_k \mid \Delta)=\Sigma_W \end{cases}$$

Further a simple computation demonstrates that Δ is multivariate normal such that

$$(2.2)\qquad \begin{cases} E(\Delta_i)=0 \\ Cov(\Delta_i) = \begin{cases} 0 & \text{if } R_i=R_t \\ 2\Sigma_B & \text{if } R_i \neq R_t \end{cases} \end{cases}$$

and

$$(2.3)\qquad Cov(\Delta_j,\Delta_k)=\begin{cases} \Sigma_B & \text{if } R_t \neq R_j \neq R_k \\ 2\Sigma_B & \text{if } R_t \neq R_k = R_j \\ 0 & \text{otherwise .} \end{cases}$$

Hence it may easily be obtained that unconditionally the joint set $z_1, \ldots z_{t-1}$ is multivariate normal such that $E(z_i)=0$ for all i and

$$\text{(2.4)} \qquad \mathrm{Cov}(z_i) = \begin{cases} 2\Sigma_W & \text{if } R_t=R_i \\ 2\Sigma_W+2\Sigma_B & \text{if } R_t\neq R_i \end{cases}$$

while

$$\text{(2.5)} \qquad \mathrm{Cov}(z_j,z_k) = \begin{cases} \Sigma_W+\Sigma_B & \text{if } R_t\neq R_j\neq R_k \\ \Sigma_W+2\Sigma_B & \text{if } R_t\neq R_j=R_k \\ \Sigma_W & \text{otherwise} \end{cases}$$

Hence the density of $z_1, \ldots, z_{t-1}$ is evaluated from (2.4) and (2.5) for any given t and any particular egg configuration and depends only on Σ_W and Σ_B the set of parameters appearing in the twin case. Estimates of Σ_W and Σ_B then are obtainable from twin data, if they are unknown.

3. A Model for the Relative Weights.

Let Y be the number of eggs at a birth that yield only like sexed offspring and assume

$$\text{(3.1)} \qquad \Pr(Y=y) = p^{y-1}(1-p) \qquad y=1, 2, \ldots,$$

(it is of course tacitly assumed that there is at least one egg at birth) where p is the relative chance of an additional egg. Further let q be the probability that an egg divides and let D be the number of divisions. Then we assume

$$\text{(3.2)} \qquad \Pr(D=d \mid Y=y) = \binom{y+d-1}{y-1} q^d (1-q)^y$$

for d=0, 1, 2, ... and y=1, 2, ... This then represents the chance that amongst y eggs there will be d divisions allowance being made for multiple divisions by every egg. Another way of viewing this is that the eggs represent urns and the offspring are distributed amongst the urns with no urn being empty. Hence the joint chance that a birth yields exactly Y eggs and D divisions is

(3.3) $$\Pr[Y=y, D=d] = \binom{y+d-1}{y-1} q^{d}(1-q)^{y}\, p^{y-1}(1-p) .$$

We shall now obtain the probability that a birth yields Y eggs given that there are a total of T (liked sexed) offspring. It is clear that T=Y+D so that for y=1, ..., t

(3.4) $$\Pr(Y=y, T=t) = \binom{t-1}{y-1} q^{t-y}(1-q)^{y}\, p^{y-1}(1-p) .$$

Summing both sides over y yields

(3.5) $$\Pr(T=t) = (1-q)(1-p)\left[q+(1-q)p\right]^{t-1} .$$

Hence dividing (3.4) by (3.5) we obtain

(3.6) $$\Pr(Y=y \mid T=t) = \binom{t-1}{y-1} \frac{q^{t-y}(1-q)^{y-1}p^{y-1}}{\left[q+(1-q)p\right]^{t-1}}$$

$$= \binom{t-1}{y-1} \frac{\gamma^{y-1}}{(1+\gamma)^{t-1}}$$

where y=1, ..., t and

(3.7) $$\gamma = \frac{(1-q)p}{q} .$$

Therefore the chance of Y eggs given T offspring depends on the single parameter γ . Consider now its interpretation: Suppose T=2 i.e., the twin case. Now for monozygotic (one egg) twins Y=1 and

(3.8) $$\Pr(Y=1 \mid T=2) = (1+\gamma)^{-1};$$

for like sexed dizygotic twins (2 eggs), Y=2 and

(3.9) $$\Pr(Y=2 \mid T=2) = \gamma/(1+\gamma)^{-1} .$$

Hence the relative frequency of like sexed dizygotic to monozygotic twins is simply γ , the ratio of (3.9) to (3.8). This quantity too then is clearly estimable from only twin data.

Now from the relations (2.4) and (2.5) and (3.6) we can determine the posterior probability that the t offspring were derived from 1, ..., t eggs. In addition for any of the values 2 to t-1, we assume that a

priori every configuration or assignment of individuals is equally likely. This then will permit us to obtain the actual posterior probability for particular individuals associated with particular eggs i.e., a complete probabilistic ascertainment of which individuals are identical twins and which are fraternal is then feasible. We then choose that case which has maximum posteriori probability. If Σ_W , Σ_B and γ are unknown, they can be estimated from twin data alone and we may choose the case that has maximum estimated posteriori probability. We shall illustrate this with triplets: Here the triplet x_1, x_2, x_3 is transformed to $x_3-x_1=z_1$ and $x_3-x_2=z_2$ and (z_1,z_2) is unconditionally multivariate normal with zero mean and covariance matrix given below for each case. Individuals within the parentheses are presumed to be from the same egg.

Case	Covariance Matrix of the Joint Distribution of (z_1, z_2)	Relative Frequency
1 egg(x_1,x_2,x_3)	$\begin{pmatrix} 2\Sigma_W & \Sigma_W \\ \Sigma_W & 2\Sigma_W \end{pmatrix}$	1
2 egg$(x_3),(x_1,x_2)$	$\begin{pmatrix} 2\Sigma_W + 2\Sigma_B & \Sigma_W + 2\Sigma_B \\ \Sigma_W + 2\Sigma_B & 2\Sigma_W + 2\Sigma_B \end{pmatrix}$	$\frac{2}{3}\gamma$
2 egg$(x_3,x_1),(x_2)$	$\begin{pmatrix} 2\Sigma_W & \Sigma_W \\ \Sigma_W & 2\Sigma_W + 2\Sigma_B \end{pmatrix}$	$\frac{2}{3}\gamma$
2 egg$(x_3,x_2),(x_1)$	$\begin{pmatrix} 2\Sigma_W + 2\Sigma_B & \Sigma_W \\ \Sigma_W & 2\Sigma_W \end{pmatrix}$	$\frac{2}{3}\gamma$
3 egg$(x_3),(x_1),(x_2)$	$\begin{pmatrix} 2\Sigma_W + 2\Sigma_B & \Sigma_W + \Sigma_B \\ \Sigma_W + \Sigma_B & 2\Sigma_W + 2\Sigma_B \end{pmatrix}$	$1-\gamma^2$

Note for the two egg case the original relative frequency is 2γ but since there are three cases which exhaust the discriminatory possibilities

we assume that they are all equally likely before hand so that the relative frequency now becomes $\frac{2}{3}\gamma$ for each of these 2 egg "subcases".

The equally likely a priori configurations for 2 egg-triplets obviously presents no difficulty. However for higher order births the same number of eggs can lead to different partitions. For example in quadruplets there are two distinct partitions in the two egg case i.e. 3 from one egg and one from the other (3,1) or two from each egg (2,2). The partition (3,1) is twice as likely as (2,2) and this must be taken into account. In general then for a fixed t and y there are $\binom{t-1}{y-1}$ compositions $(C_1, \ldots, C_y)$ for integral $C_j \geq 1$, $\Sigma C_j = t$ where C_j represents the number of the individuals belonging to j^{th} egg, as it were. We then must calculate the frequency f_p of every distinct ordered partitioned $p = (i_1, i_2, \ldots, i_y)$, subsets of the compositions, where $i_1 \geq i_2 \geq \ldots \geq i_y \geq 1$; $\sum_p f_p = \binom{t-1}{y-1}$ and the summation is over the distinct ordered partitions. Further if we define a_{pk} = number of i_j's equal to k for k=1, 2, ..., t-y+1, for a partition p then it is clear from elementary combinatorial considerations that

$$f_p = y! / \prod_{k=1}^{t-y+1} a_{pk}! \tag{3.10}$$

where $\sum_{k=1}^{t-y+1} k\, a_{pk} = t$.

Hence the a priori chance of a particular partition p of a y egg and t offspring case is

$$f_p \frac{\gamma^{y-1}}{(1+\gamma)^{t-1}} \tag{3.11}$$

Now for each distinct partition there will be b_p equally likely exhaustive assignments, the "numbered" individuals assigned to the different eggs. Hence for a particular partition p, given y and t the b_p exhaustive assignments each have prior chance

$$(3.12) \qquad \frac{f_p \gamma^{y-1}}{b_p (1+\gamma)^{t-1}} \ .$$

Elementary combinatorial analysis yields

$$(3.13) \qquad b_p = \binom{t}{i_1}\binom{t-i_1}{i_2}\binom{t-i_1-i_2}{i_3}\cdots\binom{i_y}{i_y} \Big/ \prod_{k=1}^{t-y+1} a_{pk}!$$

Hence each assignment for a given partition composed of t offspring and y eggs has prior probability

$$(3.14) \qquad y!\ \gamma^{y-1}/(1+\gamma)^{t-1}\binom{t}{i_1}\binom{t-i_1}{i_2}\binom{t-i_1-i_2}{i_3}\cdots\binom{i_y}{i_y} \ .$$

Of course first the distinct partitions must be enumerated (tables for their enumeration are abailable, see Riordon (1958,p.108) and then the assignments for each partition must also be enumerated.

In order to illustrate this point, we present the case of quadruplets arising from 2 eggs, giving the prior probabilities

Partition	Assignments	Prior Probability
(2, 2)	$(x_1,x_2),(x_3,x_4)$	$\gamma/3\ (1+\gamma)^3$
	$(x_1,x_3),(x_2,x_4)$	$\gamma/3\ (1+\gamma)^3$
	$(x_1,x_4),(x_2,x_3)$	$\gamma/3\ (1+\gamma)^3$
(1, 3)	$(x_1),(x_2,x_3,x_4)$	$\gamma/2\ (1+\gamma)^3$
	$(x_2),(x_1,x_3,x_4)$	$\gamma/2\ (1+\gamma)^3$
	$(x_3),(x_1,x_2,x_4)$	$\gamma/2\ (1+\gamma)^3$
	$(x_4),(x_1,x_2,x_3)$	$\gamma/2\ (1+\gamma)^3$

All higher order cases can be handled in precisely the same way. Hence all the tools are at hand for constructing the discriminatory apparatus in the t-tuplet case. This technique should prove useful when there are available multivariate physical measurements that are approximately normal while blood type data is either lacking or inconclusive.

References.

Richter, D. L. and Geisser, S. (1960). "A statistical model for diagnosing zygosis by ridge count." Biometrics 16, 110-114.

Riordon, J. (1958) An Introduction to Combinatory Analysis, John Wiley and Sons.